

PROCESAMIENTO DE IMÁGENES Y APRENDIZAJE DE MÁQUINA APLICADO EN DIAGNOSTICOS DE PATOLOGÍAS

IMAGE PROCESSING AND MACHINE LEARNING APPLIED TO PATHOLOGY DIAGNOSTICS

Darwin Stiven Herrán Ríos ⁽¹⁾

Santiago Rodríguez Cuellar ⁽²⁾

1. Estudiante, Universidad de Ibagué. Pasante. Semillero LUN, Ibagué, Colombia.
2420191021@estudiantesunibague.edu.co
2. GI Senagrotic Centro Agropecuario La Granja Sena Regional Tolima. Experto Línea Electrónica y Telecomunicaciones, Tecnoparque Tolima, Espinal, Colombia.
srodriguez@sen.edu.co

RESUMEN

Este Proyecto presenta la evaluación e implementación de diversas técnicas de procesamiento digital de imágenes y aprendizaje automático desarrolladas para identificar y clasificar un conjunto de imágenes adquiridas de la enfermedad conocida como neurocisticercosis en cerebros de cerdos. Estas imágenes fueron obtenidas de un proyecto de grado en el cual se mejoró la calidad e intensidad de los colores para obtener datos más consistentes y reducir la variación en las características de una imagen a otra. Se emplearon múltiples métodos de clasificación para detectar y determinar las áreas afectadas por la enfermedad en cada una de las imágenes de microscopía de cortes cerebrales. Finalmente, se llevó a cabo el proceso de evaluación de los resultados obtenidos por cada una de las técnicas implementadas, comparando su exactitud y función de pérdida, con el objetivo de determinar el método que mejor clasifica e identifica cada una de las partes de la imagen.

Palabras Claves: Fibrosis, Neurocisticercosis, aprendizaje de máquina, Segmentación y Clasificadores.

ABSTRACT

This project presents the evaluation and implementation of several digital image processing and machine learning techniques developed to identify and classify a set of acquired images of the disease known as neurocysticercosis in pig brains. These images were obtained from an undergraduate project in which the quality and intensity of the colors were improved to obtain more consistent data and reduce the variation in features from one image to another. Multiple classification methods were employed to detect and determine the areas affected by the disease in each of the brain slice microscopy images. Finally, the process of evaluating the results obtained by each of the implemented techniques was carried out, comparing their accuracy and loss function, with the aim of determining the method that best classifies and identifies each of the parts of the image.

Keywords: Fibrosis, Neurocysticercosis, Machine learning, Segmentation, Classifiers.

INTRODUCCIÓN

La medicina moderna ha experimentado una notable evolución en diversos aspectos, y uno de los factores clave en este progreso ha sido el desarrollo de nuevas herramientas tecnológicas. En particular, el procesamiento digital de imágenes se ha convertido en una rama fundamental de la ingeniería aplicada a la medicina. Esta tecnología ha permitido automatizar procesos complejos que antes requerían mucho tiempo y esfuerzo, además de generar resultados subjetivos debido a la influencia del evaluador. Gracias a las técnicas desarrolladas, los investigadores ahora cuentan con resultados más rápidos, reproducibles y confiables en comparación con los obtenidos de forma manual.

Una enfermedad que ha experimentado un notable aumento en los últimos años es la neurocisticercosis, causada por la *Taenia solium* y transmitida a través de la ingestión de alimentos contaminados con heces humanas. Esta enfermedad afecta varios órganos del cuerpo humano, siendo especialmente delicada cuando se presenta en el cerebro, donde se forman quistes cerebrales que pueden dar lugar a ceguera, convulsiones, cefaleas crónicas y epilepsia. Ante la importancia de esta enfermedad, se requiere el desarrollo de nuevos medicamentos que puedan combatirla de manera más efectiva y rápida.

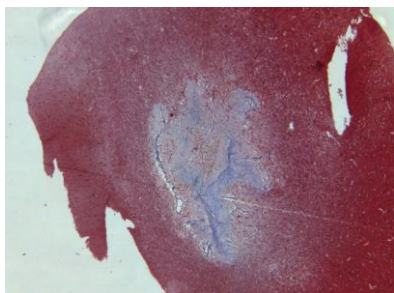


Figura 01. Fibrosis producida por neurocisticercosis en porcinos.

En el Grupo de Cisticercosis del Perú (GCP), específicamente en el laboratorio de Diagnóstico Parasitológico de la Universidad Peruana Cayetano Heredia, se están llevando a cabo investigaciones orientadas al desarrollo de nuevas alternativas para combatir esta enfermedad. Para evaluar los resultados de los nuevos productos, como anti

convulsionantes, y considerando la limitación ética de trabajar con seres humanos, se utilizan modelos animales infectados con la enfermedad como intermediarios en el estudio. Dado que la neurocisticercosis se contrae principalmente a través del consumo de carne de cerdo contaminada, se eligió a este animal como modelo para el estudio. De esta manera, los cerdos son infectados y, una vez que la enfermedad ha evolucionado, se administra el producto en estudio para evaluar sus resultados y verificar que no sea perjudicial. Para ello, se sacrifica al animal y se realizan cortes cerebrales de 5mm de espesor, los cuales se analizan en el microscopio, prestando especial atención a ciertas características de cada quiste. Uno de los aspectos relevantes en este proceso es la presencia de fibrosis alrededor del quiste, que actúa como un mecanismo de defensa del cerebro impidiendo el crecimiento o la propagación de patógenos dañinos. Para observar esta característica en las imágenes de los cortes cerebrales, se aplican tintes a los tejidos utilizando Tricrómica de Masson. Sin embargo, la detección y medición de la fibrosis en estas imágenes se realiza manualmente, lo cual es un proceso largo y subjetivo, ya que los resultados pueden variar incluso entre mediciones realizadas por el mismo investigador, y más aún entre diferentes investigadores.

Con el objetivo de mejorar este proceso se comenzó a desarrollar una solución que involucra el procesamiento digital de imágenes y el aprendizaje automático para la detección de la fibrosis en imágenes de microscopía RGB de cortes cerebrales de 5mm de espesor. Para llevar a cabo este proyecto, se realizó una búsqueda bibliográfica con el fin de recopilar información teórica y conocer los métodos existentes utilizados para detectar la fibrosis en otras áreas, como el hígado [1]. Aunque se encontraron métodos para la detección automática de fibrosis en el hígado, en este caso particular se observaron diferencias significativas entre la fibrosis y el tejido circundante, lo cual facilitaba enormemente su detección. Lamentablemente, en las imágenes de este proyecto no se presentaba esta clara distinción. También se encontraron métodos que permitían la separación de objetos en imágenes, pero requerían la intervención del usuario [2]. Dado que no existía un método adecuado para la detección de la fibrosis en imágenes de microscopía de cortes cerebrales de cerdos, se decidió aplicar distintas técnicas de detección y clasificación de

Imágenes para abordar este desafío. Se implementaron técnicas tales como el clasificador bayesiano, bayesiano con PCA, bayesiano con LDA, random forest, Adaboost, XGboost y diferentes redes neuronales, donde se analizaron y compararon los datos obtenidos en cada modelo, evaluando su precisión y función de pérdida, a fin de determinar con cual método se obtenían los mejores resultados para así poder implementarlo en estudios y aplicaciones posteriores.

ESTADO DEL ARTE

La inteligencia artificial (IA) es una rama de la informática en rápido desarrollo cuyo objetivo es diseñar y construir máquinas capaces de realizar diversas tareas cognitivas a un nivel igual o superior al de la inteligencia humana. [3–4] El aprendizaje automático (ML) es una parte de la IA que incluye algoritmos informáticos diseñados para enseñar o entrenar a una máquina a resolver un problema o realizar tareas complejas sin programación específica previa.

El aprendizaje supervisado y el aprendizaje no supervisado son dos categorías tradicionales de ML y se suelen utilizar para crear IA restringida (débil), mientras que el desarrollo de una IA general sigue siendo un objetivo a largo plazo. Hoy en día, para el análisis de datos biológicos, como los relacionados con tejidos cerebrales y enfermedades como la neurocisticercosis, existen varias herramientas y algoritmos de ML. Algunos ejemplos son los árboles de decisión (DT), las redes bayesianas (BN), los bosques aleatorios (RF) regresión logística (LR) y máquinas de vectores de soporte (SVM). Específicamente subconjunto del aprendizaje automático, denominado "aprendizaje profundo", se refiere al desarrollo de redes neuronales artificiales. (Fig.02) Las redes neuronales recurrentes y las redes neuronales convolucionales representan métodos potencialmente valiosos para el análisis de grandes datos biológicos y la identificación de patrones útiles dentro de los datos disponibles con el fin de sacar conclusiones adecuadas sobre mecanismos fisiológicos y patológicos en células y tejidos [5–6].

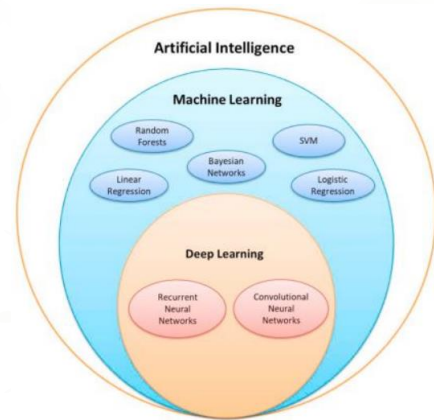


Figura 02. Inteligencia artificial, aprendizaje de máquinas y aprendizaje profundo.

El desarrollo de nuevas técnicas de adquisición y procesamiento digital de imágenes han abierto la posibilidad de extraer información cuantitativa de datos obtenidos de diferentes fuentes, entre ellas la microscopia, realizada en muestras biológicas o en la detección individual mediante propiedades cuantitativas, ayudando a aumentar drásticamente el número de preguntas que se pueden abordar experimentalmente. Estos avances van de la mano con la medicina, buscando que los resultados proporcionados por los expertos sean más exactos [7 - 8].

La medicina está constantemente en proceso de evolución, lo que implica la introducción de nuevos fármacos y terapias para tratar diversas enfermedades con el objetivo de mejorar su efectividad y eficacia. Para lograr esto, se llevan a cabo investigaciones que involucran la evaluación de diferentes medicamentos en modelos de prueba. La elección de la dirección a seguir en la investigación se basa en las respuestas que estos medicamentos muestran en relación con la enfermedad en estudio. Los métodos de procesamiento de imágenes tienen un potencial significativo para extraer información valiosa en este tipo de procedimientos y pueden utilizarse para analizar modelos animales. Un ejemplo de esto se encuentra en el estudio de los quistes generados por la neurocisticercosis en animales intermediarios, una enfermedad que está experimentando un aumento alarmante en los países en vías de desarrollo. La neurocisticercosis se caracteriza por una serie de síntomas que incluyen cefaleas crónicas, epilepsia, ceguera, convulsiones,

meningitis y demencia. Debido a la diversidad de síntomas, su diagnóstico a menudo se confunde con otras enfermedades, lo que permite que el quiste continúe creciendo y afectando al paciente, empeorando su situación. El tratamiento típico para combatir esta enfermedad consiste en la administración de un antiparasitario durante un período de un mes [9-10].

Para evaluar un medicamento es necesario conocer las características del quiste antes y durante el tratamiento. Una característica muy importante del quiste es la fibrosis aledaña a éste. Distintos trabajos se han desarrollado buscando la identificación automática de fibrosis en imágenes de diferentes partes del cuerpo, utilizando métodos de procesamiento digital de imágenes. De esta forma, se han realizado trabajos empleando un clasificador k-medias para la detección automática en el hígado de seres humanos, empleando imágenes de biopsias empleando un productor marcador para teñir la fibrosis [1], en imágenes de ultrasonido [11] y en elastografías empleando técnicas de resonancia magnética o ecografía [12]. Para obtener un resultado adecuado se emplea un protocolo de adquisición para garantizar que las imágenes tengan las mismas características, es decir que la fibrosis y cada tejido aparezcan siempre con colores o intensidades similares. La automatización requiere normalmente de clasificadores empleando descriptores de color o textura [13- 14].

RESULTADOS

A continuación, se exponen las métricas de exactitud de cada uno de los modelos implementado en la detección y clasificación de neurocisticercosis en porcinos, con el fin de poder comparar el rendimiento de todos los modelos.

Métodos de clasificación	Exactitud del modelo
Bayesiano	72.80 %
Bayesiano con PCA	77.10 %
Arboles de decisión	82.60 %
Random forest	91.30 %
XGboost	91.30 %
Adaptive boost	81.50 %

CNN	97.83%
U-net	89.04%

Tabla 01. Comparación de las métricas de exactitud por cada uno de los métodos implementados.

• Segmentación de imágenes usando un enfoque de ventana deslizante

Técnica utilizada para dividir una imagen en regiones más pequeñas llamadas ventanas y analizar cada ventana por separado. útil cuando se desea aplicar un algoritmo o modelo de aprendizaje automático a partes específicas de una imagen en lugar de la imagen completa. Se implementa esta técnica con el fin de visualizar los resultados de manera funcional de alguno de los clasificadores usados en este estudio.

Se reconstruye la imagen pixel a pixel con la predicción de cada una de las ventanas usando cada uno de los métodos entrenados y probados.

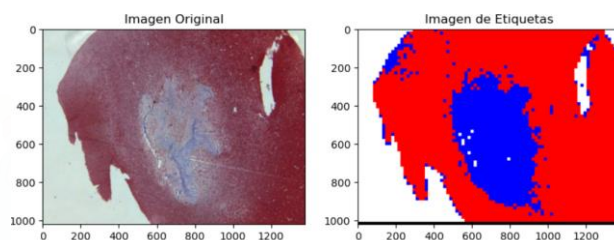


Figura 03. Segmentation de imágenes usando un enfoque de Ventana deslizante con el clasificador Random forest.

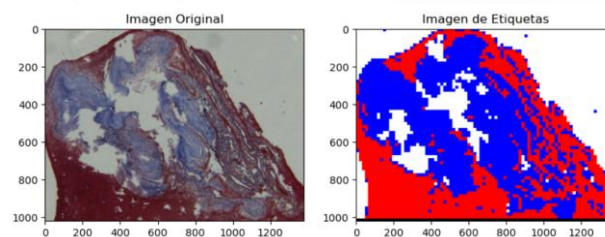


Figura 04. Segementación de imágenes usando un enfoque de Ventana deslizante con el clasificador XGboost.

Se seleccionaron dos métodos como ejemplo para la segmentación por ventanas deslizantes, a cada característica se le atribuyó un color (Azul fibrosis, rojo tejido cerebral y fondo blanco) para que al momento de que el método seleccionado clasifique cada Ventana en una clase específica se pueda reconstruir la imagen y compararla con la original.

• Segmentación con U-net

Se usaron las imágenes originales y las máscaras binarizadas para entrenar el modelo y que segmentara el tejido afectado por la fibrosis.



Figura 05. Imagen original e imagen binarizada para el entrenamiento.

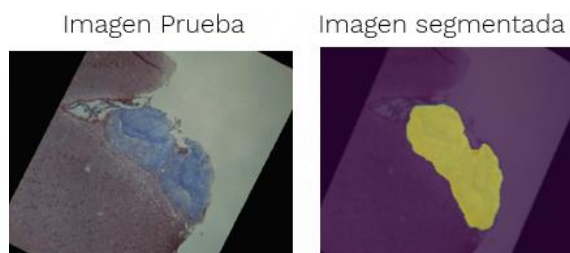


Figura 06. Imagen de prueba e imagen segmentada por la U-net.

CONCLUSIONES

1. La selección del método de clasificación más apropiado en términos de exactitud se logró a través de un proceso de prueba que involucró varios modelos distintos. Cada uno de estos modelos fue sometido a una evaluación exhaustiva para determinar su capacidad de clasificar correctamente los datos según los criterios específicos del sistema. Tras un análisis minucioso y detenido de los resultados obtenidos, se

identificó el modelo más adecuado para el propósito en cuestión. En consecuencia, se puede afirmar que la elección metódica del modelo de clasificación adecuado es esencial para garantizar un rendimiento óptimo del sistema en términos de precisión y eficiencia.

2. Gracias a la implementación de diversas técnicas de clasificación y detección de imágenes, se logró una clasificación exitosa de las áreas afectadas por la fibrosis, el tejido cerebral y el fondo en las imágenes analizadas en gran medida. Aunque hubo algunas secciones específicas en las pruebas que no se catalogaron correctamente, en general, las predicciones obtenidas en los ensayos fueron correctas.
3. Aunque los métodos clásicos de clasificación y detección, como el Bayesiano, Random Forest, XGboost y Adaboost, arrojaron resultados satisfactorios, se evidenció que, con enfoques más avanzados, como las redes neuronales, se logran tasas de exactitud y precisión más elevadas, lo que conduce a resultados superiores.

REFERENCIAS

1. Hernandez, P., Castaño, I., Travieso, C., Mompero, B., & Ortega, F. (2016). Quantification and Statistical Analysis Methods for Vessel Wall Components from Stained Images with Masson's Trichrome. *Plos One*.
2. Gallego, J. (2009). Foreground segmentation and tracking based on foreground and background modeling techniques. Universitat Politècnica de Catalunya.
3. Davidson, L., & Boland, M.R. (s.f) Towards deep phenotyping pregnancy: a systematic review on artificial intelligence and machine learning methods to improve pregnancy outcomes. *Briefings in Bioinformatics*.
4. Maffulli, N., Rodriguez, H.C., Stone, I.W., Nam, A., Song, A., Gupta, M., Alvarado, R., &

- Ramon, D. (s.f). Artificial intelligence and machine learning in orthopedic surgery: a systematic review protocol
6. Zhang, T., Zeng, Y., Zhang, Y., Zhang, X., Shi, M., Tang, L., Zhang, D., & Xu, B. (Año de publicación). Neuron type classification in rat brain based on integrative convolutional and tree-based recurrent neural networks. *Scientific Reports*.
 8. Estrada, S., et al. (2013). Neurocisticercosis. Hallazgos radiológicos. *Radiología*, 55(2), 130-141.
 9. Ford, T., Lim, D., & Mertz, J. (2012). Fast optically sectioned fluorescence, HiLo endomicroscopy. *Journal of Biomedical Optics*, 17(2).
 10. Pradilla, G. (1986). Cisticercosis: Clínica, diagnóstico y tratamiento. *Salud Uninorte*, 3(2), 115-125.
 11. Adebajo, C., et al. (2012). Ultrasound-based transient elastography for the detection of hepatic fibrosis in patients with recurrent hepatitis C virus after liver transplantation: a systematic review and meta-analysis. *Liver Transplantation*, 12, 323-331.
 12. Ebinuma, H., et al. (2011). Evaluation of liver fibrosis by transient elastography using acoustic radiation force impulse: comparison with Fibroscan. *Journal of Gastroenterology*, 46, 1238-1248.
 13. Garcia, M. (s.f). Detección y extracción de aviones en secuencias de vídeo (Tesis de licenciatura, Facultat de Matemàtiques, Universitat de Barcelona).
 14. Forero, M. (2002). Introducción al procesamiento digital de imágenes. *La Silueta Ediciones Ltda*.
 5. Alkhodari, M., & Fraiwan, L. (s.f). Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. *Computer Methods and Programs in Biomedicine*.
 7. Restrepo, B. (2000). Diagnóstico y respuesta inmune en neurocisticercosis. *Mvz – Cordoba*, 5(1), 23-25.